

## Desempenho dos candidatos da região Centro-Oeste em redações no Exame Nacional do Ensino Médio de 2017

### *Performance of candidates from the Midwest region in essays of the 2017 National High School Examination*

Dacione Sampaio Biao<sup>a</sup>, Ricardo Polato<sup>b</sup>

<sup>a</sup> *dacione.biao@gmail.com*

<sup>b</sup> *ricardopolato@hotmail.com*

**Resumo:** O Exame Nacional do Ensino Médio - Enem é uma prova realizada pelo Instituto Nacional de Estudos e Pesquisas - INEP criada em 1998 para avaliar a qualidade da educação nacional. Atualmente ela é utilizada, principalmente, como acesso ao ensino superior em diversas universidades brasileiras. O presente artigo busca através de técnicas de mineração de dados analisar informações que possam influenciar o desempenho dos participantes no resultado da prova de redação. Para isso, foi utilizada a metodologia CRISP-DM. Foi criado um modelo de árvore de decisão, com aplicação do algoritmo J48, permitindo a distinção das principais variáveis que impactam o resultado do aluno. Estas informações podem ser utilizadas para a tomada de decisão pelos gestores educacionais na formulação de estratégias relacionadas à educação brasileira.

**Palavras chave:** ENEM; Qualidade educacional; Desempenho; Mineração de dados.

**Abstract:** The National High School Examination - Enem is a test carried out by the National Institute for Studies and Research - INEP created in 1998 to measure the quality of national education. It is currently used mainly for access to higher education in several Brazilian universities. This article intends, by means data mining techniques, to analyze information that may influence the performance of participants in the result of the essay examination. For this, the CRISP-DM methodology was applied. A decision tree was built, by applying the J48 algorithm, that allows the distinction of the main variables that impact the students result. This information can be used for decision making by educational managers regarding strategies for improving the quality of Brazilian education.

**Keywords:** ENEM; Educational quality; Performance; Data mining.

## 1. Introdução

O Exame Nacional do ensino Médio (Enem) é realizado anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), desde 1998. Ele tem o objetivo de avaliar o desempenho escolar ao final da Educação Básica e colaborar para o acesso à Educação Superior, programas de bolsa de estudos e/ou de financiamento estudantil, entre outros programas do Ministério da Educação.

Um dos aspectos positivos deste exame é o favorecimento da mobilidade dos estudantes para instituições de ensino superior nos mais variados locais do país, possibilitando também que sujeitos oriundos de regiões menos desenvolvidas se desloquem para outras mais desenvolvidas. Esta mobilidade é interessante não

somente para a criação de lideranças em todos os estados da federação, mas igualmente para estabelecer um ambiente multicultural nas universidades. (Silveira et al., 2015).

O exame é constituído de quatro provas objetivas, com 45 questões de múltipla escolha cada, e uma redação em Língua Portuguesa. As provas objetivas e a redação avaliarão as seguintes áreas de conhecimento do ensino médio e os respectivos componentes curriculares:

- Linguagens, Códigos e suas Tecnologias e Redação: Língua Portuguesa, Literatura, Língua Estrangeira (Inglês ou Espanhol), Artes, Educação Física e Tecnologias da Informação e Comunicação;
- Ciências Humanas e suas Tecnologias: História, Geografia, Filosofia e Sociologia;
- Ciências da Natureza e suas Tecnologias:

Química, Física e Biologia;

- Matemática e suas Tecnologias: Matemática.

A nota de Redação é a única cuja nota pode variar de zero a mil pontos, ao passo que as provas objetivas têm uma metodologia de avaliação diferente. Trata-se da TRI - Teoria de Resposta ao Item que classifica as questões em fáceis, médias e difíceis. Com esta metodologia é avaliado o conhecimento global do estudante, não apenas a quantidade de acertos, mas a coerência entre eles.

O Inep fornece anualmente uma série de dados sobre o Enem em seu portal com objetivo desenvolver e disseminar avaliações e informações educacionais. Com estes microdados é possível acessar informações específicas, gerar análises mais aprofundadas sobre a educação brasileira por parte de pesquisadores, jornalistas e gestores públicos, por exemplo (INEP, 2019).

A presente pesquisa busca realizar uma análise exploratória da base de dados de participantes do Enem 2017 no que concerne as notas de Redação. Através da análise dos dados será possível descobrir padrões, classificar os desempenhos dos estudantes e apontar as principais variáveis que possam afetar a nota final desta prova.

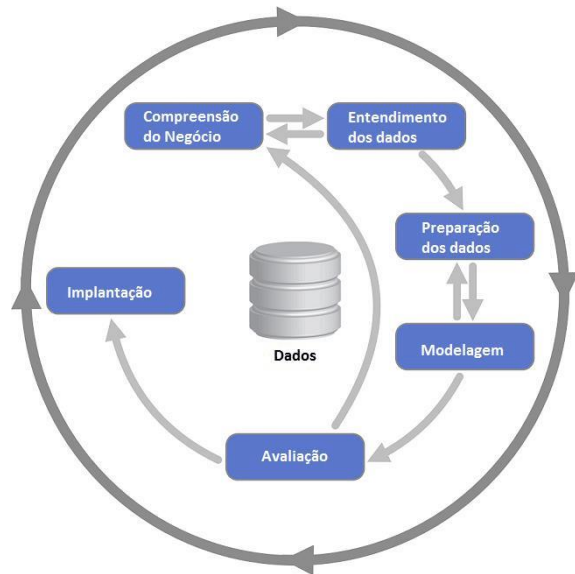
O estudo foi realizado segundo a metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*), um reconhecido framework para mineração de dados, que facilita a interpretação e descoberta de padrões.

## 2. Metodologia

Na mineração de dados há um processo que define uma estrutura para resolução do problema, permitindo consistência, repetibilidade e objetividade razoáveis, o CRISP-DM, abreviação de *Cross Industry Standard Process for Data Mining*, que pode ser traduzido como Processo Padrão Inter-Indústrias para Mineração de Dados. (Provost e Fawcett, 2013, p. 26).

O CRISP-DM é um modelo de referência que fornece uma visão geral do ciclo de vida de um projeto de mineração de dados. Ele estabelece fases e tarefas que são dependentes e se retroalimentam para atingimento aos objetivos do projeto, conforme apresentado na Figura 1.

Figura 1. Ciclo CRISP-DM.



Fonte: Chapman et al. (2001).

Ainda, segundo Provost e Fawcett (2013, p. 14), manter esse processo em mente fornece uma estrutura para concatenar o pensamento sobre problemas de análise de dados. O pensamento estruturado sobre análise enfatiza esses aspectos muitas vezes subestimados de apoiar a tomada de decisão com dados.

Segundo Wirth e Hipp (2000), este modelo é constituído por seis fases que constituem o ciclo de vida de um projeto de mineração de dados:

- **Compreensão do Negócio:** esta fase inicial concentra-se em entender os objetivos e requisitos do projeto de uma perspectiva de negócios convertendo esse conhecimento em uma definição de problema de mineração de dados e um plano de projeto preliminar projetado para atingir os objetivos.

*Como a estrutura do CRISP-DM é iterativa, a seção de entendimento de negócios será revisada frequentemente durante um grande projeto.* (Clark, 2018).

- **Compreensão dos dados:** esta fase começa com uma coleta de dados inicial e prossegue com atividades para se familiarizar com os dados, identificar problemas de qualidade de dados, descobrir os primeiros insights ou detectar subconjuntos interessantes para formar hipóteses para informações ocultas.
- **Preparação de dados:** uma vez que o cientista de dados entende qual é o caso de uso, como os dados foram coletados e outros detalhes, é necessário pré-processar os dados em uma forma utilizável para modelagem.

(CLARK, 2018). Desta forma, a fase de preparação de dados abrange todas as atividades para construir o conjunto de dados final a partir dos dados brutos iniciais. É provável que as tarefas de preparação de dados sejam executadas várias vezes e não em qualquer ordem prescrita.

- **Modelagem:** nesta fase, várias técnicas de modelagem são selecionadas e aplicadas, e seus parâmetros são calibrados para valores ótimos. Algumas técnicas exigem formatos de dados específicos e existe uma ligação estreita entre preparação de dados e modelagem. Muitas vezes, percebe-se problemas de dados durante a modelagem ou obtém-se ideias para a construção de novos dados.
- **Avaliação:** neste estágio do projeto, é constituído um ou mais modelos na perspectiva de análise de dados. Contudo, antes de prosseguir para a implantação final do modelo, é importante avaliar mais detalhadamente do modelo e revisar as etapas executadas para a construção garantindo que ele atinja adequadamente os objetivos de negócios. No final desta fase, uma decisão sobre o uso dos resultados de mineração de dados deve ser tomada.
- **Implantação:** a criação do modelo geralmente não é o final do projeto. O conhecimento adquirido precisará ser organizado e apresentado de forma que o cliente possa usá-lo. Dependendo dos requisitos, a fase de implantação pode ser tão simples quanto gerar um relatório ou tão complexa quanto implementar um processo de mineração de dados repetitivo.

Independentemente de a implantação ser bem-sucedida, o processo geralmente retorna à fase de entendimento do negócio. O processo de dados de mineração produz uma grande quantidade de informações sobre o problema de negócios e as dificuldades de sua solução. Uma segunda iteração pode gerar uma solução aprimorada. (Provost & Fawcett, 2013, p. 34).

Desta forma, o processo de mineração de dados segundo a metodologia CRISP-DM não é linear, mas sim, iterativo, onde a repetição é a regra e não a exceção. Na prática, quer dizer voltar as fases anteriores para prover insumos de melhor qualidade para as tarefas posteriores.

### 3. Aplicação da metodologia CRISP-DM

#### 3.1 Compreensão do negócio

A nota de redação pode ser um grande aliado do participante do Enem. Aqueles que zeram esta prova são eliminados da concorrência de programas do MEC como o Sistema de Seleção Unificada - SiSU, Programa Universidade para Todos - ProUni e Fundo de Financiamento Estudantil - Fies. Além disso, a nota de redação é o critério de desempate na disputa por bolsas de estudo em faculdades particulares e vagas em instituições públicas.

A redação é avaliada segundo cinco competências e em cada uma delas o estudante pode ter a nota de zero a duzentos. A nota final do participante será a média aritmética das notas totais atribuídas por dois avaliadores. As competências avaliadas são as seguintes:

- **Competência 1:** Demonstrar domínio da modalidade escrita formal da língua portuguesa.
- **Competência 2:** Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa.
- **Competência 3:** Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista.
- **Competência 4:** Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação.
- **Competência 5:** Elaborar proposta de intervenção para o problema abordado que respeite os direitos humanos.

Dada a importância do resultado da redação para o aluno que pretende ingressar em uma universidade, este trabalho tem como objetivo deste estudo é avaliar o nível das notas dos alunos para cada competência, analisando possíveis relações entre o perfil sociodemográfico dos estudantes com suas respectivas notas. Esta análise poderá ser insumos estabelecimento de políticas públicas que tragam melhoria na formação do jovem que pretende ingressar em uma universidade.

#### 3.2 Compreensão dos dados

O Portal do INEP disponibiliza os microdados

do Enem contendo em um único arquivo com informações sobre os participantes, conforme o seguinte agrupamento:

- Variáveis de controle do participante;
- Variáveis de controle da Escola;
- Pedidos de Atendimento Especializado;
- Pedidos de Atendimento Específico;
- Pedidos de Recursos Especializados e Específicos para Realização das Provas;
- Dados do Local de Aplicação da Prova;
- Variáveis das provas objetivas;
- Dados da Redação;
- Dados Questionário Socioeconômico

Nesta mesma publicação, além dos dados de cada participante, também são disponibilizadas informações das provas, gabaritos, edital, matriz de referência, desempenho dos participantes, dicionário dos dados, entre outros. Com base nas informações contidas nestes documentos foi possível aplicar técnicas de mineração de dados.

A base utilizada neste estudo é de Microdados Enem 2017, que possui 137 variáveis e registros de estudantes de todo território brasileiro, antes de qualquer pré-processamento.

Dentre os dados apresentados, está a nota de cada competência por aluno. O agrupamento desses dados por UF é apresentado na Tabela 1.

**Tabela 1.** Média de notas por UF (2017)

| UF | Competência |        |        |        |       |
|----|-------------|--------|--------|--------|-------|
|    | 1           | 2      | 3      | 4      | 5     |
| DF | 122,41      | 115,96 | 112,08 | 113,70 | 85,20 |
| GO | 118,42      | 116,12 | 109,83 | 111,32 | 83,10 |
| MS | 115,70      | 111,58 | 105,18 | 107,71 | 76,02 |
| MT | 111,19      | 108,30 | 101,20 | 103,21 | 74,47 |

**Fonte:** Resultados da pesquisa (2019).

Nota-se que o Distrito Federal tem a melhor média de notas na maioria das competências, ficando atrás de Goiás somente na Competência 2. Já o Estado com a menor média é o Mato Grosso, que fica em última posição comparado aos demais estados do Centro-Oeste. A competência 1 foi a que obteve maior média e a Competência 5 é a qual se obteve menor média em todos os Estados comparada com as demais.

Dentre os alunos do Centro-Oeste que declaram o tipo de instituição que concluiu o

ensino médio, os que concluíram no ensino regular tiveram 552,07 de média na nota final da redação. Já os que concluíram na educação especial - modalidade substitutiva alcançaram a nota de 491,77 em média e quem concluiu na modalidade de educação de jovens e adultos ficaram com nota em média de 417,41.

Dentre o total de participantes do exame no Centro-Oeste que realizaram a prova, 3.930 tiraram nota zero ou tiveram sua redação anulada. Quanto a nota máxima, apenas três alunos conseguiram alcançar os 1.000 pontos, atendendo integralmente às 5 competências.

Dos três participantes que alcançaram nota máxima, dois frequentam escola privada, um com e outro sem bolsa de estudo integral e um aluno é oriundo de escola pública.

### 3.3 Preparação dos dados

Para a preparação e pré-processamento dos dados foi utilizado Python como linguagem de programação, com a biblioteca Pandas.

Ainda como ferramentas, foram utilizadas o Anaconda Navigator, que é uma interface gráfica de usuário que permite iniciar aplicativos e gerenciar pacotes, ambientes e canais, e o Jupyter Notebook versão 5.7.8, uma ferramenta web para criação interativa de códigos que combina textos, matemática, cálculos, imagens e gráficos.

A base de dados principal foi obtida através do website do INEP por meio de um arquivo, formato csv. Desta base de dados foram selecionadas, além das notas, as informações de perfil, resultados e questionário socioeconômico dos participantes residentes em algum dos estados da região Centro-Oeste mais o Distrito Federal do Brasil.

Durante o pré-processamento, foram selecionados os inscritos no Enem 2017 que residem na região Centro Oeste, totalizando aproximadamente 562 mil registros. Ainda nesta fase, foram eliminados os inscritos que não compareceram a prova, resultando aproximadamente 355 mil registros que serviram de base para análise e mineração.

Foi observado pouca representatividade nos Indicadores de Pedidos de Atendimento Especializado, Pedidos de Atendimento Específico e Pedidos de Recursos Especializados e Específicos para Realização das Provas, sendo descartados do estudo estes atributos. As informações das provas objetivas também foram des-

cartadas, visto enfoque apenas na prova de Redação.

Foram criadas seis novas variáveis, com o objetivo de classificar o nível do participante de acordo com nota obtida em cada uma das cinco competências avaliadas, inclusive da prova de redação. A Tabela 2 detalha os conceitos atribuídos de acordo com as notas recebidas.

**Tabela 2.** Divisão por conceito.

| Nota obtida |     | Conceito |
|-------------|-----|----------|
| De          | Até |          |
| 0           | 60  | Baixo    |
| 61          | 140 | Médio    |
| 141         | 200 | Alto     |

Fonte: Resultados da pesquisa (2019).

### 3.4 Modelagem

Como informado anteriormente, o estudo não tem como objetivo inicial a construção de um modelo de alta precisão, mas sim aplicar técnicas que viabilizassem análises mais consistentes sobre as notas obtidas pelos estudantes na prova de Redação do Enem, em função do seu perfil sociodemográfico e das notas obtidas em cada competência de avaliação e que compõe a nota final de Redação.

Para aplicar as técnicas de mineração de dados foi utilizado o software o *Waikato Environment for Knowledge Analysis* (WEKA), em sua versão 3.8.1. O WEKA é um software livre criado a partir de 1993 pela Universidade de Waikato, Nova Zelândia, e tem como objetivo prover um grande número de algoritmos de aprendizado de máquina e ferramentas de processamento de dados. Sua implementação é realizada em linguagem Java, o que também contribui para sua manutenção e modificação.

A técnica utilizada foi a de árvore de decisão, gerada pelo algoritmo J48, no ambiente WEKA. J48 é uma versão do tradicional algoritmo de C4.5 que consegue, através da discretização de valores numéricos utilizando heurísticas, processar atributos de qualquer tipo na entrada. Neste algoritmo os parâmetros aprendem diretamente pelos dados. Isso é realizado através de um processo conhecido como Treinamento de modelo.

O objetivo da árvore de decisão é agrupar as instâncias de forma recursiva, minimizando a entropia (ou seja, a variabilidade das classes) dentro de cada grupo. Para isso, o algoritmo utiliza os valores de atributos em cada nó não folha e deposita as instâncias nos nós folha. Dessa forma, cada nó interno corresponde a uma decisão de classificação, e a leitura e interpretação da árvore são naturais (Rabbi, et al. 2018).

### 3.5 Avaliação

A Tabela 3 apresenta informações consolidadas da aplicação do algoritmo de mineração de dados. Observa-se que, para a Competência 1, a técnica aplicada resultou 76,07% de instâncias classificadas corretamente com índice Kappa de 0,1721, já para a Competência 2 o percentual foi de 78,37% de instâncias classificadas corretamente com Kappa de 0,0728. Na Competência 3, o Kappa chegou a 0,1079 com 77,57% de acertos. Para a Competência 4, o percentual de acurácia chegou a 78,78% com Kappa de 0,1291. Já a Competência 5 obteve uma assertividade inferior às demais com 53,74% e Kappa de 0,1234. Tão importante quanto o número de instâncias corretamente classificadas é a métrica Kappa, que indica o nível de concordância (coesão) dos dados classificados, onde sua variação vai de 0 a 1 de magnitude.

**Tabela 3.** Indicadores da Análise

| Competência | Instâncias classificadas corretamente |        | Instâncias classificadas incorretamente |        | Kappa  |
|-------------|---------------------------------------|--------|-----------------------------------------|--------|--------|
|             | Valor                                 | %      | Valor                                   | %      |        |
| 1           | 270.446                               | 76,07% | 85.091                                  | 23,93% | 0,1721 |
| 2           | 278.668                               | 78,37% | 76.889                                  | 21,63% | 0,0728 |
| 3           | 275.819                               | 77,57% | 79.738                                  | 22,43% | 0,1079 |
| 4           | 280.094                               | 78,78% | 75.463                                  | 21,22% | 0,1291 |
| 5           | 191.067                               | 53,74% | 164.490                                 | 46,26% | 0,1234 |

Fonte: Resultados da pesquisa (2019).

Além destes indicadores, a matriz de confusão contém informações importantes para o entendimento do resultado do algoritmo, dentre elas: a quantidade de instâncias classificadas corretamente; a quantidade de instâncias classificadas erroneamente; a quantidade de instâncias que o algoritmo acreditava ser de um tipo e na verdade foram classificadas como outro. A Tabela 4 representa as matrizes de confusão para cada Competência analisada.

**Tabela 4.** Matrizes de confusão relativas às cinco competências ( $C_i$ ).

|       | Real<br>Predito | a       | b      | c     |
|-------|-----------------|---------|--------|-------|
|       |                 |         |        |       |
| $C_1$ | a               | 256.343 | 10.804 | 48    |
|       | b               | 49.765  | 14.120 | 32    |
|       | c               | 24.252  | 190    | 3     |
| $C_2$ | a               | 274.182 | 4.462  | 27    |
|       | b               | 37.661  | 4.484  | 16    |
|       | c               | 34.636  | 87     | 2     |
| $C_3$ | a               | 268.781 | 5.906  | 42    |
|       | b               | 36.551  | 7.037  | 26    |
|       | c               | 37.072  | 141    | 1     |
| $C_4$ | a               | 272.012 | 6.470  | 40    |
|       | b               | 38.155  | 8.081  | 22    |
|       | c               | 30.670  | 106    | 1     |
| $C_5$ | a               | 126.839 | 56.408 | 3.801 |
|       | b               | 76.822  | 62.558 | 1.592 |
|       | c               | 21.601  | 4.266  | 1.670 |

Fonte: Resultados da pesquisa (2019).

Na Competência 1, o algoritmo previu corretamente 256.343 instâncias do conceito Médio (a). Entretanto, previu incorretamente 10.804 para o conceito Alto (b) e 48 para Baixo (c). Para o conceito Alto, foi previsto corretamente 14.120 contra 49.797 instâncias incorretas e para o conceito Baixo foram 3 corretas e 24.442 incorretas.

Quanto a competência 2, o algoritmo previu corretamente 274.182 instâncias contra 4.489 erros no conceito Médio (a). No conceito Baixo (b) obteve 4.484 acertos contra 36.677 erros. Já o conceito Baixo (c) que obteve uma previsão correta de apenas 2 contra 34.723 previsões incorretas.

Na competência 3, foi obtido um total de 275.819 previsões corretas contra 79.738 incorretas, considerando os três conceitos.

A competência 4 apresentou uma Matriz de Confusão com 280.094 previsões assertivas e 75.463 previsões erradas entre os três conceitos, com maior assertividade no conceito Médio (c).

A última competência apresentou um total de 191.067 acertos, contra 164.490 previsões incorretas. Esta foi a competência com menor percentual de previsões corretas, o que a torna de baixa previsibilidade para a análise.

Como resultados, as variáveis que apresentaram forte influência para a classificação de cada competência foram:

- Competência 1. Demonstrar domínio da modalidade escrita formal da língua portuguesa: (i) Tipo de escola frequentou o Ensino Médio; (ii) Sexo do participante; (iii) Idade do participante; (iv) Situação de conclusão do Ensino Médio; (v) Série de estudo do pai ou homem responsável;
- Competência 2. Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa: (i) Tipo de escola frequentou o Ensino Médio; (ii) Idade do participante; (iii) Série de estudo da mãe ou mulher responsável; (iv) Situação de conclusão do Ensino Médio; (v) Sexo do participante.
- Competência 3. Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista: (i) Tipo de escola frequentou o Ensino Médio; (ii) Idade do participante; (iii) Série de estudo da mãe ou mulher responsável; (iv) Ano que concluiu o Ensino Médio; (v) Sexo do participante.
- Competência 4. Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação: (i) Tipo de escola frequentou o Ensino Médio; (ii) Idade do participante; (iii) Situação de conclusão do Ensino Médio; (iv) Residência possui telefone; (v) Cor ou raça do participante.
- Competência 5 - Elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos: (i) Tipo de escola frequentou o Ensino Médio; (ii) Idade do participante; (iii) Acesso à Internet em sua residência; (iv) Cor ou raça do participante; (v) Residência possui computador; (vi) Situ-

ação de conclusão do Ensino Médio.

Com base nas Árvore de Decisão criadas para cada Competência, nota-se como variáveis importantes recorrentes e que influenciam no resultado da redação: o tipo da escola em que frequentou o Ensino Médio, a idade, sexo do participante e a situação de conclusão do Ensino Médio.

Na Figura 2, são apresentados gráficos comparando de cada uma das variáveis com a média da soma das notas dos cinco conceitos avaliados. Soma esta que representa a nota final da redação atribuída ao participante.

O primeiro gráfico apresenta a média da nota final da redação por tipo de escola frequentada pelo aluno durante o Ensino Médio. Através deste, podemos verificar que o grupo de alunos que frequentou o Ensino Médio somente em escola privada sem utilização bolsa (Item D) obteve uma média superior aos demais, atingindo nota 667,56. Seguido pelo grupo de alunos que frequentou o Ensino Médio somente em escola privada com utilização de bolsa (Item E) com nota 629,81. Em seguida, temos os alunos que frequentaram o Ensino Médio parte em escola pública e parte em privada com bolsa (Item D), com nota 550,76, os que frequentaram o Ensino Médio parte em escola

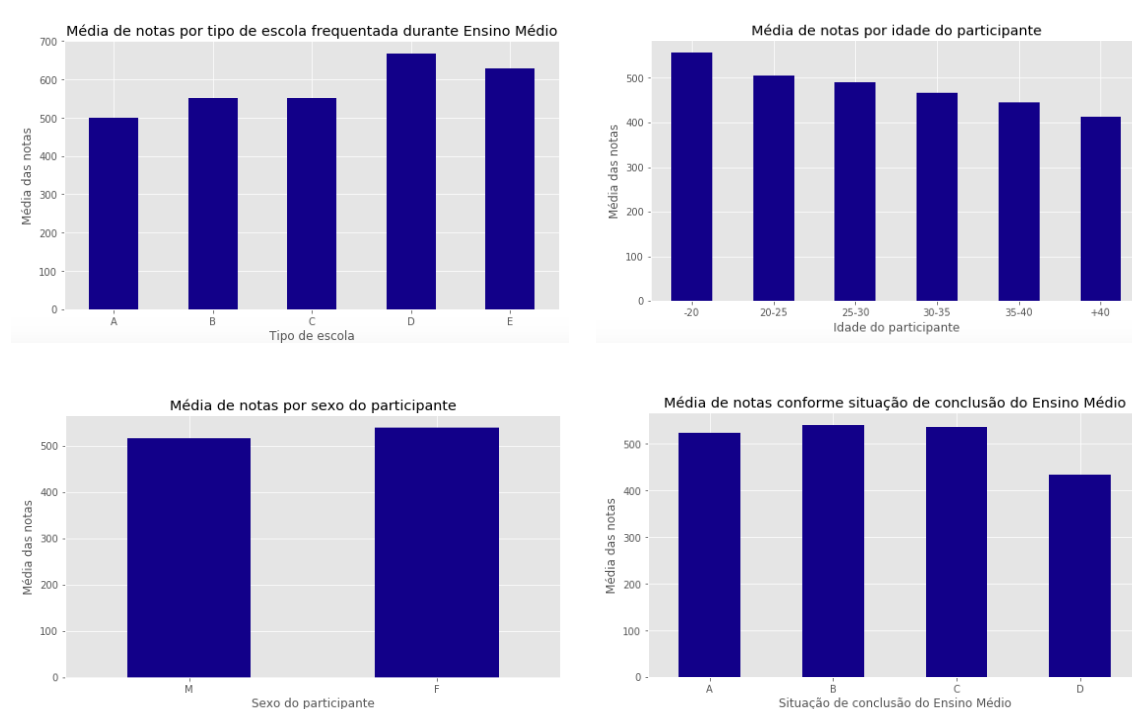
pública e parte em privada sem bolsa (Item B) com nota 550,72. E, o grupo de alunos que frequentaram o Ensino Médio somente em escola pública (Item A) atingiu a menor média de notas, ficando com 498,53.

O segundo gráfico traz informação da média da nota final da redação por idade do participante. Analisando este gráfico, nota-se que quanto maior a idade do participante, menor é sua média na redação onde, alunos com idade menor que 20 anos obtiveram média de 555,63, alunos entre 20 e 25 anos obtiveram média de 504,59, 25 a 30 anos, média 489,01, 30 a 35 anos tiveram média de 467,23, 35 a 40 anos, 445,26 de média e alunos com idade superior a 40 anos obtiveram uma média de 413,59 de média na avaliação da redação.

Quanto ao terceiro gráfico, expressa a média da nota final da redação conforme o sexo do participante. Nota-se que as alunas obtiveram média de 538,93, enquanto os participantes do sexo masculino ficaram com média de 515,56.

O quarto gráfico apresenta informações da média da nota final comparando com situação de conclusão do Ensino Médio. O grupo de alunos que atingiu a maior média foram os que estão cursando e concluirão ainda no ano 2017 (Item B), ficando com média de 539,70. Em

**Figura 2.** Comparação das variáveis com a média da soma das notas dos cinco conceitos avaliados.



Fonte: Resultados da pesquisa (2019).

seguida, os alunos que estão cursando e concluirão o Ensino Médio após 2017 (Item C), com média de 535,48. O grupo de alunos que já concluiu o Ensino Médio (Item A) ficou com média de 523,36 e os que não concluíram e não estão cursando (Item D) ficaram com 433,25 de média na redação.

#### 4. Conclusão

Neste trabalho buscou a utilização de técnicas de mineração de dados para encontrar padrões e classificar o perfil dos participantes do Enem residentes na região Centro-Oeste do Brasil e através deste analisar as variáveis que impactam o desempenho na nota da redação.

A partir dos resultados extraídos, foi evidenciado que o tipo da escola em que o aluno frequenta o Ensino Médio, a idade, sexo do participante e a situação de conclusão do Ensino Médio são fatores que interfere na qualidade da redação produzida durante o exame.

Como principal quesito que interfere no resultado é tipo de escola, sendo que alunos que se utilizam da educação pública são diretamente afetados obtendo uma nota inferior aos

demais. Desta forma, a educação pública do Brasil no que tange ao Ensino Médio ainda necessita ser melhorada, tanto politicamente quando pedagogicamente, e as classes sociais mais baixas são a mais afetadas, influenciando diretamente no desempenho do estudante.

Outro quesito importante é quanto a idade do aluno e o tempo que se passou desde que terminou o Ensino Médio até a realização do exame. Isto torna importante o estímulo e a orientação ao aluno quanto a continuidade dos estudos. Professores e equipe pedagógica devem orientar a busca de uma vocação, definindo áreas de interesse para a inserção do aluno no Ensino Superior pois quanto antes for realizado o exame, maior a probabilidade de um melhor resultado.

Para trabalhos futuros, poderão ser utilizadas técnicas de mineração de dados aplicadas aos resultados do Enem por escola, identificando as quais obtiveram melhores resultados e os quesitos que estas melhores escolas possuem a fim de propor adoção destes quesitos para as demais escolas com notas inferiores e, dessa forma para que possam melhorar o desempenho de seus alunos.

#### Referências

- Almhadeen, L., Akkaya, M., & Sari, A. Mining Student Data Using CRISP-DM Model. *International Journal of Computer Science and Information Security*. (IJCSIS), Vol. 15, No. 2, February 2017.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: step-bystep data mining guide*, CRISP-DM Consortium. <ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf>.
- Clark, A. *The Machine Learning Audit: CRISP-DM Framework*. ISACA Journal Volume 1, 2018.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP. *Redação no ENEM 2018: cartilha do participante*. Brasília: INEP/DAEB, 2018.
- Provost, F. & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic tinkering*. Sebastopol: O'Reilly Media.
- Rabbi, B., Rabbi, D. B. K., Gonçalves, V. S.; Gonçalves Júnior, E. R.; Brasil, J. A. (2018). Mineração de dados aplicada a base de reclamações sobre produtos e serviços do programa de proteção e defesa do consumidor. *Brazilian Journal of Development*, 4(5).
- Silveira, F. L., Barbosa, M. C. B. & SILVA, R. (2015). Exame Nacional do Ensino Médio (ENEM): Uma análise crítica. *Revista Brasileira de Ensino de Física*, 37(1).
- Wirth, R. & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*.

#### Sobre os autores

##### *Dacione Sampaio Biao*

Funcionária do Banco do Brasil desde 2007, é bacharel em Administração pela Universidade Federal da Bahia (2002), especialista em Negócios Financeiros pela Universidade do Vale do Rio dos Sinos (2011), especialista em Gerenciamento de Projetos de TI pela AVM (2016), e mestranda em Governança, Tecnologia e Inovação pela Universidade Católica de Brasília.



***Ricardo Polato***

Assessor de Tecnologia no Banco do Brasil, é bacharel em Administração pelo Centro de Ensino Superior de Rondonópolis - CESUR (2005), Tecnólogo em Análise e Desenvolvimento de Sistemas pelo Centro Universitário da Grande Dourados (UNIGRAN), especialista em Gestão Pública Municipal pela Universidade de Brasília - UNB (2019) e mestrando em Governança, Tecnologia e Inovação pela Universidade Católica de Brasília - UCB.